DEPARTMENT OF NATIONAL DEFENCE
CANADA

OPERATIONAL RESEARCH DIVISION

DIRECTORATE OF OPERATIONAL RESEARCH
(CORPORATE, AIR, & MARITIME)
&
DIRECTORATE OF STRATEGIC HUMAN RESOURCE CO-ORDINATION

ORD PROJECT REPORT PR0004

A NEW MODEL FOR PREDICTING CF MOVING COSTS

by

Aline Chouinard

JANUARY 2000

OTTAWA, CANADA

National Defence Défense nationale

OPERATIONAL RESEARCH DIVISION

CATEGORIES OF PUBLICATION

ORD Reports are the most authoritative and most carefully considered publications of the DGOR scientific community. They normally embody the results of major research activities or are significant works of lasting value or provide a comprehensive view on major defence research initiatives. ORD Reports are approved personally by DGOR, and are subject to peer review.

ORD Project Reports record the analysis and results of studies conducted for specific sponsors. This Category is the main vehicle to report completed research to the sponsors and may also describe a significant milestone in ongoing work. They are approved by DGOR and are subject to peer review. They are released initially to sponsors and may, with sponsor approval, be released to other agencies having an interest in the material.

Directorate Research Notes are issued by directorates. They are intended to outline, develop or document proposals, ideas, analysis or models which do not warrant more formal publication. They may record development work done in support of sponsored projects which could be applied elsewhere in the future. As such they help serve as the corporate scientific memory of the directorates.

ORD Journal Reprints provide readily available copies of articles published with DGOR approval, by OR researchers in learned journals, open technical publications, proceedings, etc.

ORD Contractor Reports document research done under contract of DGOR agencies by industrial concerns, universities, consultants, other government departments or agencies, etc. The scientific content is the responsibility of the originator but has been reviewed by the scientific authority for the contract and approved for release by DGOR.

DEPARTMENT OF NATIONAL DEFENCE
CANADA

OPERATIONAL RESEARCH DIVISION

DIRECTORATE OF OPERATIONAL RESEARCH
(CORPORATE, AIR, & MARITIME)
&
DIRECTORATE OF STRATEGIC HUMAN RESOURCE CO-ORDINATION

ORD PROJECT REPORT PR0004

A NEW MODEL FOR PREDICTING CF MOVING COSTS

by

Aline Chouinard

Recommended by: _____ DOR(CAM) Approved by: A Bradfield DGOR

ORD Project Reports present the considered results of project analyses to sponsors and interested agencies. They do not necessarily represent the official views of the Canadian Department of National Defence

OTTAWA, CANADA JANUARY 2000

102-02252/

ABSTRACT

This paper presents the first version of the Canadian Forces Cost Moves prediction model. The model can predict the cost of moves when basic information about the particulars of a move are supplied. This first version of the model is limited to domestic moves (i.e. within Canada). The model is a linear regression model.

RÉSUMÉ

Ce rapport présente la version initiale du modèle des coûts de déménagement des Forces Canadiennes. Le modèle peut prédire les coûts de déménagement à partir des données de base d'un déménagement. Cette première version du modèle se limite aux déménagements au Canada. Le modèle est un modèle de régression linéaire.

TABLE OF CONTENTS

ABSTRACT/RÉSUMÉ i
TABLE OF CONTENTS ii
EXECUTIVE SUMMARY iv
I. INTRODUCTION 1
AIM 1
II METHODOLOGY 2
III CHOICE OF PREDICTORS 3
IV. DATA PREPARATION 3
INDEPENDENT VARIABLES 3
Missing Data 5
COST COMPONENTS 6
INFLATION 6
SPECIAL CASE OF THE WEIGHT OF FURNITURE AND EFFECTS 6
V. ANALYSIS OF COVARIANCE 7
DISCUSSION I 8
Cost Components 8
Independent Variables 10
Interactions 11
VI. THE GENERAL LINEAR MODEL 11
DISCUSSION II 11
Inaccurate Data 11
Lack of observations 11
Non-linearity and indirect measures 12
VII. GENERATING PREDICTIONS 13
VIII FUTURE WORK 14
IRPP ADJUSTMENTS 14
FOREIGN MOVES 14
FURTHER ISSUES 14
IX. CONCLUSION 15
X. RECOMMENDATIONS 16
LIST OF REFERENCES 17

ANNEX A LIST OF LOCATIONS AND MOC GROUPINGS..... A-1
ANNEX B MARITAL STATUS VERSUS REAL ESTATE FEES/HOUSE HUNTING TRIP COSTS..... B-1
ANNEX C THE HOUSE HUNTING TRIP REGRESSION MODEL..... C-1

EXECUTIVE SUMMARY

This study has been requested by the Directorate of Military Career and Relocation Management (DMCARM). A model for predicting the cost of moves will be a valuable tool in planning the budget of proposed posting plots, and thereby avoid large budget overruns. This is the second report in the study. The first report provided estimates of cost moves to be used on an interim basis. Prior to that the client was using a single overall cost moves average to predict the cost of moves. This report describes the development of a new tool for predicting CF moving costs. This model provides a moving cost estimate for each CF member appearing on the posting plot.

The total predicted cost for the cost of a move is the sum of the individual cost components. Cost components are: transportation, interim lodgings and meals, movement of furniture and effects, etc. There are forty such costs.

A model was built for each cost component. The resulting models are linear equations of the form:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

where "y" is the cost component to be predicted (the dependent variable), $x_1, x_2, \ldots, x_k$, are the variables which are good predictors (independent variables), and $b_0, b_1, b_2, \ldots, b_k$ are numerical coefficients that have been determined by regression analysis.

The predictor variables for each cost components are taken from the following list: distance of the move, the weight of furniture and effects, the rank of the CF member, the location to, the location from, the marital status, the number of dependants, the occupation and the years of service.

Prediction models (i.e. regression models) are based on historical data. Given the CF's extensive experience in moving its personnel, the moving costs of similar cases from previous years can be "averaged" to provide an estimate for the cost of proposed moves. The data for this study came primarily from the Cost Moves Master File (CMMF), which was a database supported by the Directorate of Human Resources Information Management (DHRIM) until March 98. Following this, the database is no longer supported. The following manipulations were made to the database in order to proceed

with this study: the foreign moves were taken out of the database, the number of cost components was reduced to thirteen by lumping some costs together, the marital status and years of service were merged from another database, etc.

The special case of the weight of furniture and effects is noteworthy. The data collected on the weight of furniture and effects in the CMMF were sparse. A regression model was built to estimate this variable using the available sparse data. Then the predicted weight of furniture and effects was used to fill in missing data for that field.

Next, analysis of covariance is done to decide which variables are good predictors for each cost component. The regression analysis is performed next and linear equations are established for each cost component.

Once the linear models are built for each cost component, the model can generate predictions by substituting the appropriate values for the independent variables into the linear equations. In fact, it also provides confidence limits around the estimate as well. Then predicted inflation is added to the estimates for the coming year.

This completes Phase II of this study. The next phase of this study will address the following issues: to adjust the model to predict costs under the new integrated relocation pilot program (IRPP), a model for foreign moves, etc.

This new model for predicting CF moving costs can be a valuable tool to control the cost move budget. It can provide more accurate estimates than any tools used so far by DMCARM. It is also recommended that a cost move database be created and maintained to assist the cost moves prediction process. The quality of the data must be improved and be readily compatible with the model.

A NEW MODEL FOR PREDICTING CF MOVING COSTS

I. INTRODUCTION

1. This is the second report in the ongoing Cost Moves study. This study has been requested by the Directorate of Military Career and Relocation Management (DMCARM). A considerable amount of work was done on Cost Moves by the Directorate of Manpower Analysis in the 80's. See for example, references 1-13. The 80's model grouped the moves into 28 separate subgroups and derived an estimate for each subgroup. In 1989, another model was developed, using a regression model. This model never made it as a production model used by the client. In the 90's, the client was simply calculating one overall cost move average as an estimate for the cost of moves.

2. The first report completed in the current study (Ref.14) provided estimates of Cost Moves to be used on an interim basis. It presented the various phases that have to be completed for this study. It also presented an overview of the model envisioned.

3. A cost moves model will be useful in planning the budget of proposed posting plots, and thereby avoid large budget overruns. The annual cost moves budget is around $160 Million.

AIM

4. This report presents the latest Cost Moves Prediction tool. It can estimate the cost of domestic moves, that is a move that takes place between two Canadian locations.

II. METHODOLOGY

5. The total predicted cost for the cost of a move is the sum of the individual cost components. A general linear model was built for each cost component. The resulting model is a linear equation of the form:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

where "y" is the cost component to be predicted (the dependent variable), $x_1, x_2, \ldots, x_k$, are the variables which are good predictors (independent variables), and $b_0, b_1, b_2, \ldots, b_k$ are numerical coefficients that have been determined by regression analysis. In the case of a non-numerical independent variable such as rank, the corresponding term in the linear model is b_{ki}. The coefficient b_{ki} takes on a different value matching the ith value of the independent variable (k refers to the kth independent variable). This is called a mixed model.

6. It was decided to build a model for each individual cost component rather than develop a model for the total cost of a move. In this way, the estimates for each cost component will be available to include in the cost formula of the new Integrated Relocation Pilot Project (IRPP)(Ref. 15).

7. The following steps have been completed in building this model:

a) choosing the variables to be tested;
b) preparing the data;
c) performing the analysis of covariance for each cost component;
d) building the linear model; and
e) generating predictions.

8. The statistical software package SAS has been used to conduct this study. The analysis of covariance will determine which independent variables best predict the individual cost components. The R^2 value will often be referred to in this report. It is a measure of how strong the linear relationship is between the independent variables and the dependant variable. R^2 equals one when points lie on a line.

III. CHOICE OF PREDICTORS

9. A choice of the list of variables to be tested as good predictors had to be made. These variables have to be readily available on some database or other. For instance, the value of the house of a member at the time of a move would be a good predictor for the cost of real estate fees but these data are not available on the databases kept by the department. The list of variables that were selected is as follows: location from, location to, rank, marital status, occupation, number of dependents, years of service, distance of the move and weight of furniture and effects (F&E).

IV. DATA PREPARATION

10. The Cost Moves Master File (CMMF) is a database which can hold up to ten years of data on Cost Moves. While this database is not supported any more by the Directorate of Human Resources Information Management (DHRIM), the author of this study managed to obtain a copy of this "extinct" database. The data available for this analysis covers the period from April 89 to March 98.

11. The data fields that are contained in the CMMF are: move number, service number, rank, occupation, date, unit moved from, unit moved to, weight of F&E, distance, move category, resource code, total cost, previous move number and the individual cost components. There are forty cost components such as the cost of transportation of the CF member and his/her dependants, the cost of interim lodgings and meals, the cost of moving F&E, and so on. The CMMF does not include the costs for the Guarantied Home Sale Program, but includes the Home Equity Assistance Program.

12. Domestic moves represented 60% of the moves in the database.

INDEPENDENT VARIABLES

13. Some variables which would potentially be good predictors of costs are not available in the CMMF (Ref.14). These variables are: marital status, years of service, family size and distance moved. The location from and location to fields have to be

identified by communities instead of the unit code, because there are many unit codes per community.

14. Data from the Military Personnel Database were added (merged) to the CMMF with regards to marital status and years of service. When merging the marital status, the merge was effected using the year of the record and the service number as keys. Some records were left without marital status. As a second step, those "orphaned" records were merged with the previous year of the record and the service number. The merge of the years of service worked well.

15. The number of dependants was derived from the CMMF itself as it is not available any other way. The first two columns of Table 1 show the definition of category as it appears in the CMMF. The third column is the number of dependants that was assumed for this study. This approach, although not ideal, was preferable to ignoring this variable altogether. This is one aspect of future data collection that has to be emphasized, the number of dependants should be specified for future development.

TABLE I
NUMBER OF DEPENDANTS

Category	Definition	Number of Dependants
0	Relocation from basic training (TQ3); takes precedence over all other categories	unknown
1	Serviceman only (no dependants)	0
2	14 or fewer years of service, and 1 or 2 dependants	1.5
3	15 or more years of service and 1 or 2 dependants	1.5
4	14 or fewer years of service and 3 or 4 dependants	3.5
5	15 or more years of service and 3 or 4 dependants	3.5
7	5 or more dependants	5

16. The database contains the Unit Identification Codes (UIC) for the location from and the location to fields. The UICs were first converted to the name of the communities where these are located, using the UIC addresses. Initially, the database contained some 200 locations (from and to). These had to be reduced to a more manageable number, yet

be sufficiently large to realistically model the domestic moves of the CF. At Ref. 16, CF locations had been judicially grouped in order to develop a representative sample of the CF population for the purpose of drug testing. The same set of locations (Tables 4 and 2 of Ref. 16) were used. After screening the foreign locations from the database, the 43 domestic locations chosen were sufficient to account for 91% of the records. These locations are listed at Annex A.

17. The distance field of the database was found to be missing considerable data. A 43 by 43 distance table was compiled using the Canadian Distance Guide (Ref. 17). This table was used to fill the distance field.

18. There were some 150 valid military occupations, these had to be reduced to a more manageable number. They were reduced to ten groups as shown at Annex A. The new occupations and generals do not have moves recorded in our database and as such do not appear in the database. The new occupations will be treated in the model like their counterparts in the old system.

Missing Data

19. There were difficulties with missing data. The domestic moves dataset contains 98,627 records, out of which a total of 36,916 had some data missing. Table II shows the breakdown of the missing data. Note that some records had more than one missing field. The records with missing 'number of dependants' were those records that were identified as category 0. The records that have distance missing are those records that had locations other than the 43 locations retained. The records with missing years of service or marital status are records for which no matches were found in the personnel database. The records with missing or invalid rank originate from the CMMF itself, they listed "unknown" as the rank.

20. When data are missing in the independent variables, SAS ignores the entire record. The records with missing data were screened out of the dataset.

COST COMPONENTS

21. Some costs components were combined in the cost "other". These are the cost components that individually amounted to less than 1% of the total expenses in the CMMF. Examples of such cost components would be movement of pet (e.g. horses) or movement of a boat.

TABLE II
MISSING VALUES

Variable	Number of affected records	Percentage of total
Dependants	13,907	14.1%
Distance	9,787	9.9%
Years of Service	1,799	1.8%
Marital Status	16,248	16.5%
Rank	6,181	6.3%

22. The cost components 'legal costs' and 'purchase lawyer fees' were grouped together as it is unclear what the differences are between those two costs.

INFLATION

23. All costs were then converted using the average annual inflation rates (Ref. 18) to 98/99 dollars. A single overall inflation rate per year was used for all cost components.

SPECIAL CASE OF THE WEIGHT OF F&E

24. Since the field for the weight of furniture and effects had considerable missing data, it was of limited usefulness in its initial state. However, there were sufficient data to build a model to estimate the weight of furniture and effects. Analysis of covariance on the weight of furniture and effects yielded that the significant factors in estimating the weight of furniture and effects were: years of service, rank, marital status and number of dependants. A linear regression model was built with those independent variables to estimate the weight of furniture and effects. For each record of the database, the field for the weight of furniture and effects was completed using this model. Thus the weight of F&E becomes available to use as an independent variable in our main model.

25. If this had not been done, there were two alternatives. The weight of F&E could have been rejected as a variable, and the model would have used years of service, rank, marital status and number of dependants as indirect measures in its place. Or, if the weight of furniture and effects had been retained, the number of usable records by SAS would have been very small. Replacing the weight of F&E using a linear model turned out to be a very creative means of circumventing these issues.

V. ANALYSIS OF COVARIANCE

26. The next step is to perform an analysis of covariance to decide which variables are good predictors for each cost components. For each cost component analysed, a separate set of data was created containing only records for which this cost was greater than zero. Extreme points were also screened out. The screening criterion used was: costs greater than 3 standard deviations away from either side of the mean were screened out. Thirteen sets of data (datasets) were created, one for each cost component. Table III shows the number of observations in the datasets after all screening is done.

27. The analysis of covariance could be performed in three different ways:

a) All independent variables at once;

b) A forward/backward approach, in which the independent variable are added/deleted one at a time until an optimum model is reached such that as few independent variables as possible are used while optimising the R^2 value. In other words, only the variables that contribute the most to the R^2 are retained and the variables that add little to the R^2 are discarded; and

c) The analysis of covariance is performed for each year of data separately.

28. For large datasets, the method in a) sometimes yields that all the variables are significant, in which case the method in b) will provide more parsimonious and credible models. For the method in c), the number of observations for some of the costs can be small and no factors are found to be significant and the results for the ten years of data have to be compared with one another. The method in b) is a multi-step method and can be time intensive and the analyst has to decide when variables add "little" to the model. The analysis was done using all three methods, and the results were compared.

29. All three methods were applied to the thirteen datasets and for the most part, the three procedures produced similar results. The GLM procedure in SAS was used to do these analysis. Table IV shows the results. The factors are: location from (UICF), location to (UICT), maritai status (MSTATUS), rank, occupation (MOC), dependants (DEPNDTS), years of service (YOS), distance and weight of furniture and effects (weight of F&E). The blanks indicate that the factor was not significant in modelling the corresponding cost component.

TABLE III

NUMBER OF NON-ZERO OBSERVATIONS/

COMPLETE RECORDS/ NO OUTLIERS

Cost Component	Number of non-zero observations
Movement of Furniture & Effects	40,252
Interim Lodgings and Meals	39,287
Transportation	43,557
Real Estate Fees	10,327
House Hunting Trip	24,453
Movement Grant	39,580
Purchase Disbursements	12,068
Miscellaneous Legal Fees	18,913
Mortgage Default Insurance	6,517
Unaccompanied Baggage	2,595
Home Equity Assistance	805
Long Term Storage	1,416
Other	25,607

DISCUSSION I

Cost Components

30. For the movement of furniture and effects, distance and weight fully describe this cost as expected.

31. For interim lodgings and meals, the number of dependants and the distance of the move are the main predictors. Distance provides an indirect measure of the duration of

he interim lodgings; the farther away is the move, the longer it will take to get the furniture and effects into the new home. The analysis also reveals that the interim lodgings are as likely to take place in the community of origin as in the community of arrival.

TABLE IV

ANALYSIS OF COVARIANCE RESULTS

Cost Component	UICF	UICT	Rank	Mstatus	MOC	DEPNDTS	YOS	Distance	Weight of F&E
Movement of Furniture and Effects								Y	Y
Interim Lodgings and Meals	Y	Y				Y		Y	
Transportation	Y	Y				Y		Y	
Real Estate Fees	Y		Y	Y			Y		
House Hunting Trip		Y		Y				Y	
Movement Grant									
Purchase Disbursements		Y	Y						
Misc Legal Fees	Y	Y	Y	Y					
Mortgage Default Insurance		Y	Y	Y					
Unaccompanied Baggage	Y	Y							
Home Equity Assistance	Y		Y						Y
Long Term Storage		Y							
Other	Y	Y	Y	Y					

32. For transportation, the main predictor is distance, but is insufficient to describe the model for this cost. When the number of dependants, the location to and location from are added, the model is improved.

33. For real estate fees, the main predictors are location from and the rank. The rank is directly linked to the salary of the member and impacts on the size of home a member is likely to own. Similarly, the marital status is relevant as a married spouse may be earning a second income and the family may in turn own a larger home. Years of service also is significant. Not only does the member earn more (within his/her rank) but the member may have accumulated assets over time and may progressively be buying larger homes.

34. For the house hunting trip, the predictors are distance, location to and marital status as expected. The marital status probably has an effect on the duration of the house hunting trip, as it may take longer to find a home that will satisfy two persons versus one.

This assertion was verified by tabulating the average cost of the house hunting trip versus the marital status and is shown at Annex B.

35. For the movement grant, it was found that a linear model does not represent this cost adequately.

36. The purchase disbursements are probably proportional to the value of the home purchased and the best predictors of the value of a home are its location and the income of its owner. It is no surprise that the predictor variables for this cost are the location to and the rank. The same comments apply to the home equity assistance cost, the mortgage default insurance cost and to the miscellaneous legal fees cost.

37. The predictors for the long term storage cost are the weight of the furniture and effects and the location to, which appears reasonable.

38. The predictors for the unaccompanied baggage costs are the locations from and to which also appears quite reasonable.

Independent Variables

39. It appears that the number of dependants does not affect many of the costs speculated initially (Ref.14). There could be two reasons for this. First, it could be that the data for the number of dependants are too inaccurate for any relationship to surface. Or, the number of dependants really does not affect the size of the home that CF members buy. The size of the home bought depends on what one can afford and family size could be a secondary factor. If in the future, the number of dependants becomes a reliable piece of information, the relationship between costs and the number of dependants could then be assessed differently.

40. The occupation is not significant to any of the cost components. When analysis is performed without the location variables, the occupation then becomes significant. It then serves as a substitute for the locations from and to, because members from an occupation tend to have postings involving the same locations. For example, hard Navy occupations (officers) will move between Halifax, Esquimalt and Ottawa.

41. The years of service variable is not significant for predicting most costs except the real estate fees and only marginally for that component.

Interactions

42. Interactions among the factors were not tested in this phase of the study. They could be tested in a later phase of the study, and their analysis could lead to improved model accuracy.

VI. THE GENERAL LINEAR MODEL(GLM)

43. Table V shows the R^2 values obtained for these linear models.

44. An example of a linear model is provided at Annex C. The case of the house hunting trip was selected for this example.

DISCUSSION II

Inaccurate Data

45. The R^2 values appear to be low, however, given such a massive amount of data and the large number of levels for each of the numerous independent variables, *most* of those values are reasonable. The data points were from nine years of cost move data, during which some policies have changed and spending patterns have evolved. All these could contribute to spreading the data points and thereby lowering the R^2 values. Inaccuracies in the data and inaccuracies in the inflation rates are also problematic. A model can only be as good as the data used to produce it.

Lack of Observations

46. Although nine years of data were available for the analysis, for some of the cost components, there may not have been as many data points as would be desirable for building a linear model, see Table III. Costs falling in this category are: the home equity assistance program, unaccompanied baggage, long term storage and mortgage default insurance.

Non-Linearity and Indirect Measures

47. The cost components with the lowest R^2 are either not well represented by a linear model or would be better described by variables not included in the current list of predictors.

TABLE V
R^2 VALUES OF THE REGRESSION MODELS

Cost Component	R^2
Movement of Furniture & Effects	0.44
Interim Lodgings and Meals	0.12
Transportation	0.62
Real Estate Fees	0.28
House Hunting Trip	0.34
Movement Grant	0.02
Purchase Disbursements	0.33
Miscellaneous Legal Fees	0.17
Mortgage Default Insurance	0.14
Unaccompanied Baggage	0.11
Home Equity Assistance	0.20
Long Term Storage	0.12
Other	0.05

48. The mortgage grant cost is certainly not well represented by a linear model. In fact, it is a function, that takes on only two values. If the weight of furniture and effects exceeds 2000 pounds, the grant is set at a fixed value around $800 and if the weight of furniture and effects is less than 2000 pounds the grant is set at around $250. This clearly is not a linear function.

49. The home equity assistance program is a cost component that is not well represented by the variables available for the analysis. The following variables would be more appropriate: the value of the house *at the time of purchase* and the value of the house *at the time of sale*. The variables 'location from' and rank only give an indirect

measurement of the value of the house, it makes no distinction between its value at the time of sale nor at the time of purchase.

50. It is not surprising to find low R^2 value for cost components that are described using indirect measures. The case of the movement of furniture and effects illustrates that when a component is represented by *direct* measures like distance and the weight of the furniture and effects, the R^2 is higher.

51. Future model development will be aimed at exploring additional possible predictor variables and testing them in the model and hopefully thereby increasing the R^2 values.

VII. GENERATING PREDICTIONS

52. The next step is to use the GLM to predict the cost of proposed moves. Predicted values for the cost of moves can be obtained by substituting the appropriate values for the independent variable into the linear equation. There is one caveat, the categorical independent variables can only take on the values that are within the range of the model. The possible values for the locations are listed at Annex A. The marital status, rank, occupation need to be filled with valid entries. The numerical variables (the number of dependants, years of service, distance and weight of furniture and effects) can take on any non-negative value.

53. A table of the proposed moves is entered into the model and the model returns the predicted cost in an extra column along with a confidence interval. In fact, the model produces an estimate/confidence interval for each cost component. The cost components estimates are added up to provide a total cost move estimate. Then estimated inflation is added to the estimates for the coming year.

VIII. FUTURE WORK

IRPP ADJUSTMENTS

54. To adjust the estimates to account for the new relocation pilot program is not a simple task because the new policy is substantially different from the old one. Under the new policy, CF members will have the opportunity to be more involved in their moves and may even be able to collect some cash in the process. This involves a cultural change for the CF and there will be a period of adjustment, and as such moving costs are going to be very difficult to predict. Prediction models are based on historical information and as long as we have not gathered sufficient data under the new system, a prediction model cannot be built. Nevertheless, it is in times of uncertainty that they are most needed.

55. For the time being, two estimates could be provided. The first estimate is produced by the model based on historical data collected under the old system. The second will be based on the compensation formula for the new program. Since each cost component is estimated separately in the model, the estimates obtained can be "plugged into" the IRPP compensation formula to obtain an IRPP cost move estimate.

FOREIGN MOVES

56. A parallel model has to be developed for foreign moves. Foreign moves need to be analysed separately in order to produce a more effective model both for the domestic and the foreign moves.

FURTHER ISSUES

57. The model needs to become user friendly, so that it can be handed over to the client as a tool to generate cost move estimates as required. There are major hurdles before getting to that stage. Since the CMMF database is "extinct", the cost moves data will have to be obtained by some other means. The data collected on the cost of moves should contain all the independent variables required by the model and be put into a format which will be compatible with the format used by the model.

58. The current model does not make use of the information on the previous move of the CF member. If the CF moved the member in the past, the weight of furniture and effects of the member is then known prior to moving him/her. Other information from previous moves could lead to further model enhancements.

59. Investigating new independent variables and interactions among the independent variables could both be the topic of further research.

IX- CONCLUSION

60. This report describes in detail the analysis that was performed to develop the prototype version of the new Cost Moves Prediction model. Modelling the Cost of Moves for the CF was a challenging endeavour. The deliverable in this project is a tool to predict the cost of moves, a very concrete product. This model will be put to the test time and time again. The data in this analysis had to be put together from various sources and screened for quality. This takes a long time and is tedious. The analysis itself is sophisticated. It involves artful selection of variables of both types, numerical and categorical. Also, separate models were developed for each cost rather than one model for the overall cost, thirteen times more work. This new cost moves prediction model could be a valuable budget control tool for the Department.

X. RECOMMENDATIONS

61. It is recommended that the new prediction model for CF moving costs be used by the Department for planning its cost moves budget. It is further recommended that a database of cost moves be created and maintained to assist in the prediction of moving costs.

62. Specifically, the cost moves database should contain the following information for each move: the service number, the rank, the occupation, the years of service, the date of the move, the move category, the location from (rather than UIC), the location to, the weight of F&E, the distance, the marital status, the number of children and each of the cost components (clearly defined).

63. Other information that would be useful to collect would be: the value of the house sold, the value of the house purchased, the duration of interim lodgings and meals and the duration of the house hunting trip. The previous move number should be collected as it could be used in future model enhancements.

LIST OF REFERENCES

1. Project Report PR192, "The Calculation of Cost Move Commitment and Control System Averages from the Cost Move Master File", by R. Rodgers and J. Hudson, October 1982.

2. D Man A Staff Note SN 1/83, "Updated Calculations Of Cost Move Commitment Control System Averages From the Cost Move Master File", by R.Rodgers, July 1983.

3. Project Report PR239, "Solution of the Base Year Requirement Segment of the Long Term Cost Move Planning System for Other Ranks", by R. Rodgers, February 1984.

4. D Man A Staff Note SN 3/84, "An Alternate Method for the Solution of the Base Year Requirement Segment of the Long Term Cost Move Planning System for Other Ranks", by R. Rodgers, July 1984.

5. D Man A Staff Note SN 9/84, "How Significant is More? An Analysis of the Cost Move Increase in Weight of Furniture and Effects", by R. Rodgers, September 1984

6. Project Report PR251, "Calculations of Fiscal Year 1984/1985 Cost Move Commitment Control System Averages From the Cost Move Master File", by R. Rodgers, August 1984.

7. Project Report PR374, "Fiscal Year 1986/1987 Cost Move Commitment Control System Averages Calculated From the Cost Move Master File", by R. Rodgers, July 1986.

8. Project Report PR423, "Annual Calculation of the Cost Move Commitment Control System Recommended Averages For Fiscal Year 1987/88", by G. Harris, May 1987

9. D Man A Staff Note SN 1/88, "Modifications to the cost move commitment control system and estimates for FY 88/89", by A. Chouinard, April 1988.

10 D Man A Staff Note SN 5/89, "Calculation of the Cost Move Commitment Control System Projected Averages for Fiscal Year 1989/90", by S. Isbrandt and G. Harris, May 1989.

11. D Man A Staff Note SN 89/2, "Framework for a New Cost Moves Model: Analysis of Factors and Estimation Model Design", by P. Comeau, December 1989.

12. D Man A Staff Note SN 9/89, "An Approach for Estimating Average Cost of Moving Military Personnel To, From, and Between Locations Outside of Canada", by Ian Mack and G. Brenholen, November 1989.

13. D Man A Staff Note 3/91, "The Development of a New Cost Move Model", by Liam Keast and A. Chouinard, August 1991.

14. DOR(CAM) Research Note RN9906, "Cost Moves Estimates", by A. Chouinard and Sumeet Pelia, April 99.

15. Personnel Newsletter, February 99.

16. ORA Project Report PR635, "The Design Of The Survey On Illicit Drug Use In The Canadian Forces, Operation Cascade II", by A. Chouinard, February 93.

17. Canadian Distance Guide, a reference available on the Banyan Network.

18. 1998 Economic Model and Historical Rates Model, Directorate of Managerial Accounting & Comptrollership (DMAC).

ANNEX A
DOR(CAM) PROJECT REPORT PR0004
JANUARY 2000

LIST OF LOCATIONS AND MOC GROUPINGS

1. The list of the 43 domestic locations follows:

Aldergrove	Bagotville
Borden	Calgary
Chatham	Chilliwack
Cold Lake	Comox
Cornwall	Debert
Edmonton	Gagetown
Gander	Goosebay
Greenwood	Halifax
Kingston	London
Masset	Millcove
Moncton	Montreal
Moose Jaw	Newport Corner
North Bay	Ottawa
Penhold	Petawawa
Portage	Quebec
Shelburne	Shilo
St-Jean	St-John's
Suffield	Summerside
Toronto	Trenton
Vancouver	Victoria
Wainwright	Winnipeg
Yellowknife	

2. The grouping of the occupation is as follows (the name in the SAS code is in brakets):
officers in land operations (ofopsl); 21, 22, 23;
officers in air operations (ofopsa); 31, 32, 39, 63, 64, 65;

officers in sea operations (ofopss); 71;
officers in engineering occupations (ofeng); 41, 42, 43, 44, 45, 46, 24;
officers in other occupations (ofother); 51, 52, 53, 55, 56, 57, 58, 61, 62, 66, 67, 68, 69, 72, 73, 74, 75, 77, 98;
NCM's in land operations (ncmopsl); 011, 021, 022, 031, 041, 042;
NCM's in air operations (ncmopsa); 081, 091, 121, 131, 161, 168, 171, 511, 512, 513, 514, 521, 524, 525, 526, 531, 551, 561, 562, 563, 565, 572;
NCM's in sea operations (ncmopss); 065, 181, 191, 262, 273, 274, 275, 276, 283, 284, 285, 286, 312, 313, 314, 321, 331, 332, 341, 342;
NCM's in maintenance occupations (ncmeng); 052, 111, 141, 142, 151, 211, 212, 225, 226, 227, 291, 411, 421, 431, 432, 433, 435, 441, 490, 541, 611, 612, 613, 614, 615, 621, 622, 623, 624, 625, 631, 651, 811; and,
NCM's in support occupations (ncmsup); 711, 713, 714, 715, 716, 718, 719, 722, 724, 725, 731, 831, 841, 861, 862, 871, 881, 911, 921, 933, 935.

ANNEX B
DOR(CAM) PROJECT REPORT PR0004
JANUARY 2000

MARITAL STATUS VERSUS REAL ESTATE FEES/HOUSE HUNTING TRIP COSTS

1. The next table shows the average real estate per marital status, the averages are markedly higher for the married and the widowed members. It is interesting to observe that common-law members have real estate fees closer to those of single members, rather that those of married members. It seems that only one of the common-law partners invests in real estate.

TABLE B-I
MEAN REAL ESTATE FEES VERSUS MARITAL STATUS

Marital status	Average Real Estate fees
common-law involving two CF members	6426.93
common-law relationship	6535.60
married	7354.94
married to another CF member	7766.83
separated	6608.63
divorced	6240.04
single	6261.69
widowed	7820.89

2. Table B-II shows the average house hunting costs per marital status, the averages are markedly higher for the married/common-law members. It is interesting to see that the average for the widowed members lies between the averages for the single/separated/divorced and the married/ common-law members.

TABLE B-II

MEAN HOUSE HUNTING TRIP VERSUS MARITAL STATUS

Marital status	Average House Hunting Trip Costs
common-law involving two CF members	2623.73
common-law relationship	2353.96
married	2351.59
married to another CF member	2508.39
separated	1552.53
divorced	1667.80
single	1667.80
widowed	2008.20

ANNEX C
DOR(CAM) PROJECT REPORT PR0004
JANUARY 2000

THE HOUSE HUNTING TRIP REGRESSION MODEL

1. The estimate for the house hunting trip cost is given by:

$$Y=2055.85+(0.5762*\text{distance})+\text{location to factor}+\text{marital status factor}$$

The marital status factor is given in Table C-1 and the location to factor is given in Table C-2.

2. The following example serves to illustrate how this equation is used. Suppose a married member moves from Moose Jaw to Winnipeg. The distance between Moose Jaw and Winnipeg is 403 miles (Ref.3). The location to factor corresponding to Winnipeg is (-502.02). The marital status factor for a married member is (236.49). So the estimate for the house hunting trip for this member is $2022.52.

3. Note that Tables C-1 and C-2 are valid only for the house hunting trip costs. Tables like these have been derived individually for each cost listed in the report. Note that the cost factor corresponding to Yellowknife is 0 and that the cost factors for all the other locations are negative. SAS uses as the baseline the last item on the list of the categorical data, all the other items on the list are assessed in contrast to this item. Yellowknife happens to be a very far away location and no doubt a house hunting trip to Yellowknife will cost more than a house hunting trip to any other location. Places like Aldergrove, Goose Bay, Masset or Mill Cove, probably offer inexpensive lodgings as well as limited choice of home to choose from, and the duration of the house hunting trip will likely be short.

TABLE C-1

MARITAL STATUS FACTORS

FOR HOUSE HUNTING TRIP

Marital Status	factor
Common-Law (Reg F)	562.92
Common-Law Relationship	238.05
Divorced	-488.09
Married	236.49
Married (Reg F)	347.80
Separated	-575.37
Single	-329.15
Widowed	0.00

TABLE C-2
LOCATION TO FACTOR
FOR HOUSE HUNTING TRIP

Location	factor	Location	factor
Aldergrove	- 1007.83	Bagotville	-463.78
Borden	-774.95	Calgary	-610.07
Chatham	-621.45	Chilliwack	-743.54
Cold Lake	-638.33	Comox	-676.32
Cornwall	-964.76	Debert	-878.45
Edmonton	-628.51	Gagetown	-493.66
Gander	-355.01	Goose Bay	-1416.27
Greenwood	-807.94	Halifax	-766.98
Kingston	-741.31	London	-625.40
Masset	-2420.15	Mill Cove	-1216.77
Moncton	-619.23	Montreal	-821.37
Moose Jaw	-521.52	North Bay	-500.78
Ottawa	-737.49	Penhold	-951.97
Petawawa	-712.83	Portage	-816.68
Quebec	-653.73	Shelburn	-765.77
Shilo	-681.89	St-John's	-639.94
St-Jean	-784.93	Suffield	-578.18
Summerside	-447.98	Toronto	-806.28
Trenton	-764.91	Vancouver	-873.74
Victoria	-981.89	Wainwright	-785.01
Winnipeg	-502.02	Yellowknife	0.00

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM
(highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA
(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared e.g. Establishment Sponsoring a contractor's report, or tasking agency, are entered in Section 8).

PERSONNEL OPERATIONAL RESEARCH TEAM, ADM(HR-Mil), NATIONAL DEFENCE HQ, OTTAWA, CANADA, K1A 0K2

2. SECURITY CLASSIFICATION (overall security classification of the document, including special warning terms if applicable)

Unclassified

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title)

A New Model For Predicting CF Moving Costs

4. AUTHORS (last name, first name, middle initial)

Chouinard Aline

5. DATE OF PUBLICATION (month Year of Publication of document)

November 1999

6a. NO OF PAGES (total containing information. Include Annexes, Appendices, etc.) 25

6b. NO OF REFS (total cited in document) 18

7. DESCRIPTIVE NOTES (the category of document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Project Report

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address).

Director of Military Career and Relocation Management

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

NIL

9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written.)

NIL

10a. ORIGINATOR's document number (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

ORD PR0004

10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification.)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors: further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

This paper presents the first version of the Canadian Forces Cost Moves prediction model. The model can predict the cost of moves when basic information about the particulars of a move are supplied. This first version of the model is limited to domestic moves (i.e. within Canada). The model is a linear regression model.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified . If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Prediction of CF Moving Costs
Analysis of Covariance
Regression Model